
Hilltop Simulation: Passaic River RI/FS

Table of Contents

Algorithm Description	1
Diagnostic Checking	3
Statistical Distribution of Synthetic Data	7
Plan View Maps Illustrating Synthetic Surfaces	9
Begin Simulation Loop	10
Draw a sample from each cell	10
Plan View of a few Synthetic Maps	11
Actual vs Predicted Surface	13
Calculate Actual and Predicted Post Removal SWAC	16
Area weighted average -- Revised to accomodate unequal decision unit sizes	16
Actual vs Predicted SWAC	17
Distribution of Ratio of Actual to Predicted SWAC	18
Actual vs Predicted Remedial Footprint Area	19

This program simulates the accuracy of the hilltopping algorithm using one sample per potential removal area. The program simulates a large number of synthetic maps with three zones of contamination levels similar to those observed in left shoal, right shoal and channel of the Lower Passaic River in the vicinity of river mile 9 to 11.

Algorithm Description

Following is a short description of the simulation algorithm

1. Simulate mean zero spatially correlated residual process $X(x,y)$
2. Add the log-scale zone means to each of three separate zones
3. Exponentiate synthetic log-scale map
4. Subdivide each strip into 8 rectangular decision units (24 total units)
5. Select one sample from each of the 24 units (zsamp)
6. Identify those samples exceeding the selected action limit (RAL)
7. *Predicted SWAC* =mean of sample with removed samples set to zero
8. *Actual SWAC* =mean of DU synthetic surface after setting to zero
9. Calculate ratio (R) of Actual to Predicted SWAC
10. Repeat nsim > 1000 times and summarize results graphically

```
clear all  
close all  
  
rng(121199788) % Initialize random number generator
```

```
UseReportSettings =1

if UseReportSettings ==0 % Leave these settings fixed to reproduce results
    % in the report
nsim=1000;

RAL=500;
mul=6.7;
mu2=4.6;
mu3=6.7;
s2=1.6; % Variance of log-transformed data

a=1/64; % 1/a is the long flow range of influence (64 feet)
% Note that the effective range of influence is
% (1/a)*pi or 201 feet when a=64.
% the covariance model is exp(-h/h0) where h0 is the
% ellipsoidal base of the spectral density model.
%
b=a*5; % 1/b is the cross flow range of influence which also
% has effective range of influence (1/b)*pi or
% pi*64/5=40.2 feet

theta=90*pi/180;
nugget=0.0 % Nugget as a proportion of the sill variance
model='exp';

n1=64;
n2=512;

dx1=5;
dx2=5;

% Generate geographic coordinates for simulation grid
[XI, YI]=meshgrid(0:dx2:(n2-1)*dx2,0:dx1:(2*n1-1)*dx1);

alpha=0;
beta=1;
param=[alpha, beta]; % Dummy variables ignored when model ='exp'

else % Modify these parameters to experiment with other combinations
    % of Simulation Parameters
nsim=1000;

RAL=500;
mul=6.7;
mu2=4.6;
mu3=6.7;
s2=1.6; % Variance of log-transformed data

a=1/102; % 1/a/pi long flow range of influence
b=a*5; % 1/b/pi cross flow range of influence
theta=90*pi/180;
nugget=0.3 % Nugget as a proportion of the sill variance
```

```
model='exp';

n1=64;
n2=512;

dx1=5;
dx2=5;

[XI, YI]=meshgrid(0:dx2:(n2-1)*dx2,0:dx1:(2*n1-1)*dx1);
alpha=0;
beta=1;
param=[alpha, beta]; % Dummy variables ignored when model ='exp'
end

[X,S]=simfast1(2*n1,n2,dx1,dx2,param,theta,a,b,s2,model);

UseReportSettings =
1

nugget =
0.3000
```

Diagnostic Checking

```
T1=2*n1*dx1;
T2=n2*dx2;

df1=1/T1;
df2=1/T2;

sum(S(:))*df1*df2

% Define A to be the complex series of Fourier Amplitudes derived from the
% 2-D Fourier Transform of the synthetic contaminant map.

A=dx1*dx2*fft2(X);
% Borgman FFT Conventions (Equation 18.4 Borgman 1973,
% Statistical Properties of Fast Fourier Transform
% Coefficients Computed From Real Valued, Covariance
% Stationary, Periodic Random Functions. University of
% Wyoming College of Commerce and Industry, Research
% Paper Series No. 23.

% Check that synthetic data have the specified spectral density by
% calculating spectrum of simulated surface

Shat=(real(A).^2+imag(A).^2)/T1/T2;
```

```
% Calculate the covariance function for the synthetic surface using the
% FFT of the spectrum because covariance and spectrum are Fourier pairs.

Chat=real(df1*df2*fft2(Shat));

disp('Make sure simulation reproduces specified variance.')
disp(['Area under estimated 2-D spectrum is the variance: ' num2str(sum(Shat(:)))]
disp(['Should also match variance of synthetic surface: ' num2str(var(X(:)))]))

disp('Check for match to covariance function')

% Calcualte the theoretical covariance function from the spectral denwity
% model
C=real(df1*df2*fft2(S));

% Plot spectrum
figure
subplot(2,1,1)
pcolor(XI,YI,fftshift(Shat)), shading('flat')
axis('equal')
title('Spectral Density')
colorbar

subplot(2,1,2)
pcolor(XI,YI,fftshift(Chat)), shading('flat')
axis('equal')
title('Covariance Function')
colorbar

figure

plot(XI(1,1:60),Chat(1,1:60)/1.6,'-.','color',[0.8,0.8,0.8],'Linewidth',3)
hold on
plot(XI(1,1:60),C(1,1:60)/1.6,'b-')

hold on
plot(YI(1:30,1),Chat(1:30,1)/1.6,'--','color',[0.8,0.8,0.8],'Linewidth',3)
plot(YI(1:30,1),C(1:30,1)/1.6,'r-')
grid on

plot(get(gca,'xlim'),exp([-pi -pi]),'k--')

legend('Long-Flow Correlation Function: Simulated',...
    'Long-Flow Correlation Function: Theoretical',...
    'Cross Flow Correlation Function Simulated',...
    'Cross Flow Correlation Function Theoretical')

xlabel('Distance Between Points (ft)')
ylabel('Correlation (Unitless)')
```

Hilltop Simulation:
Passaic River RI/FS

```
title('Long- and Cross-flow Correlation Models')
text(130,0.66,'Long-flow Model: C(h) = e^{-h/64}','fontsize',10)
text(130,0.6,'Cross-flow Model: C(h) = e^{-h/12.8}','fontsize',10)
text(130,0.54,['Long-flow Range: ' num2str(round(1/a*pi)), ' feet.'],...
    'fontsize',10)
text(130,0.48,['Cross-flow Range: ' num2str(round(1/b*pi)), ' feet.'],...
    'fontsize',10)
text(130,0.42, ['Mean of Simulated Surface: ',num2str(mean(X(:)))])
text(130,0.36, ['Variance of Simulated Surface: ',num2str(var(X(:)))])
% Subsample and use the upper half of the simulated array to eliminate
% periodic wrap around from left to right bank edges

X=X(1:n1,1:n2);

M=[mu1*ones(20,512);mu2*ones(24,512);mu3*ones(20,512)];
```

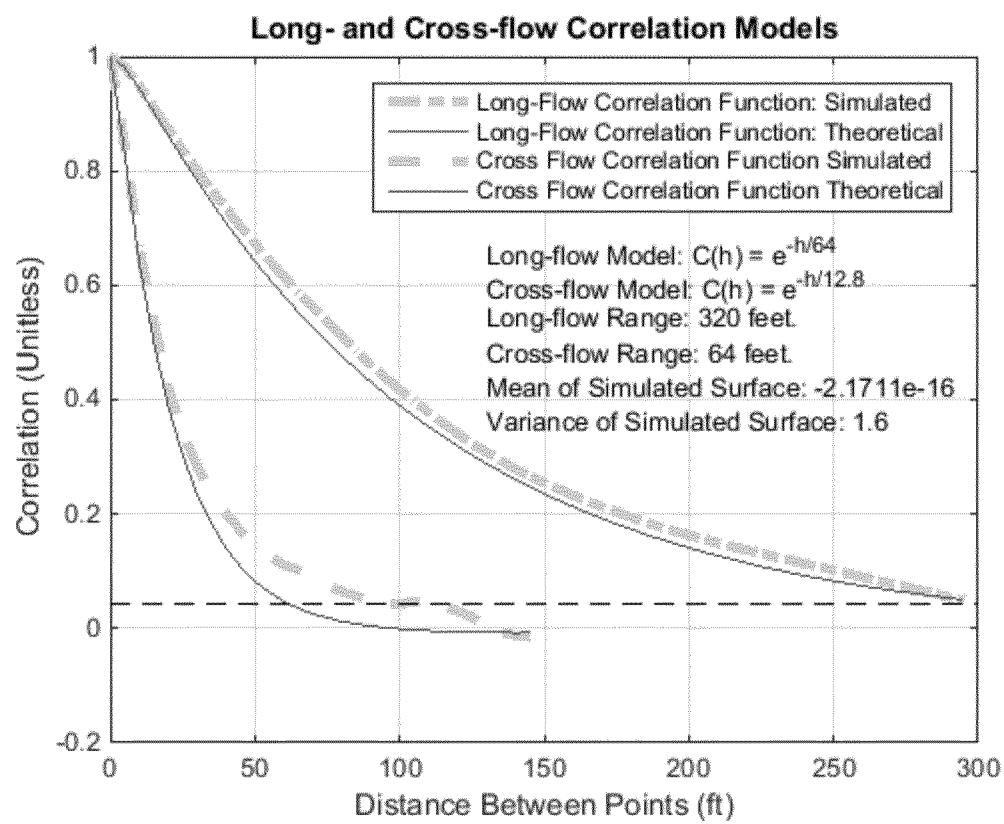
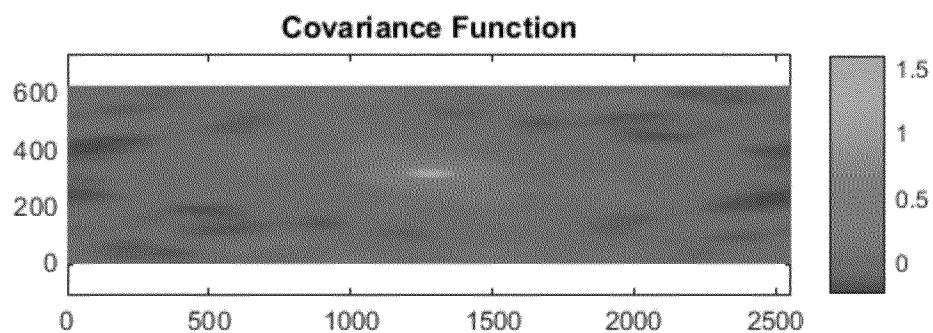
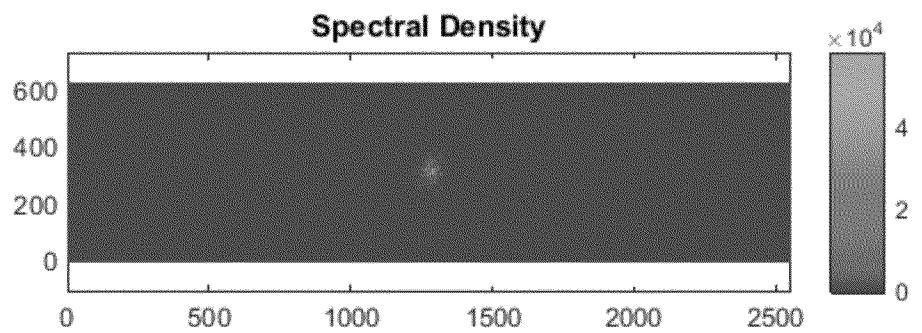


```
if UseReportSettings ==1
    Z=exp(X+M);
else
    Z=exp(X+M+nugget*random('Normal',0,sqrt(s2),size(X)));
end
```



```
ans =
1.6000
```

*Make sure simulation reproduces specified variance.
Area under estimated 2-D spectrum is the variance: 1.6
Should also match variance of synthetic surface: 1.6
Check for match to covariance function*



Statistical Distribution of Synthetic Data

The synthetic distributions are generally similar to 2378 TCDD concentrations found along the lower Passaic River. The distribution includes lower concentrations in the channel and higher concentrations on the left and right shoals.|

```

StripID=[1*ones(20,512);2*ones(24,512);3*ones(20,512)];

figure
j=find(StripID==1);
h=cdfplot(Z(j));
set(h,'color','r')
hold on

j=find(StripID==2);
h=cdfplot(Z(j));
set(h,'color','b','linestyle','--')
hold on

j=find(StripID==3);
h=cdfplot(Z(j));
set(h,'color','k');
set(gca,'xscale','log');

legend('Channel','Left Shoal','Right Shoal','location','NorthWest');
xlabel('Concentration');
ylabel('Probability');
set(gca,'xticklabel',(get(gca,'xtick')));

figure
set(gcf,'position',[433    408    484    156]);
rnames={'Geometric Mean','Mean','75th Percentile','Standard Deviation'};
colnames={'Left Shoal','Channel','Right Shoal'};

data=[round([geomean(Z(StripID==1)), geomean(Z(StripID==2)) geomean( Z(StripID==3))
round([mean(Z(StripID==1)), mean(Z(StripID==2)) mean( Z(StripID==3))])
round([prctile(Z(StripID==1), 75), prctile(Z(StripID==2),75) prctile( Z(StripID==3),75))]
round([std(Z(StripID==1)), std(Z(StripID==2)) std( Z(StripID==3))])];

t=uitable(gcf,'Data',data,'ColumnName',colnames,'Position',[40 40 425 100], 'RowName',rnames);

% Define 24 Decision Units

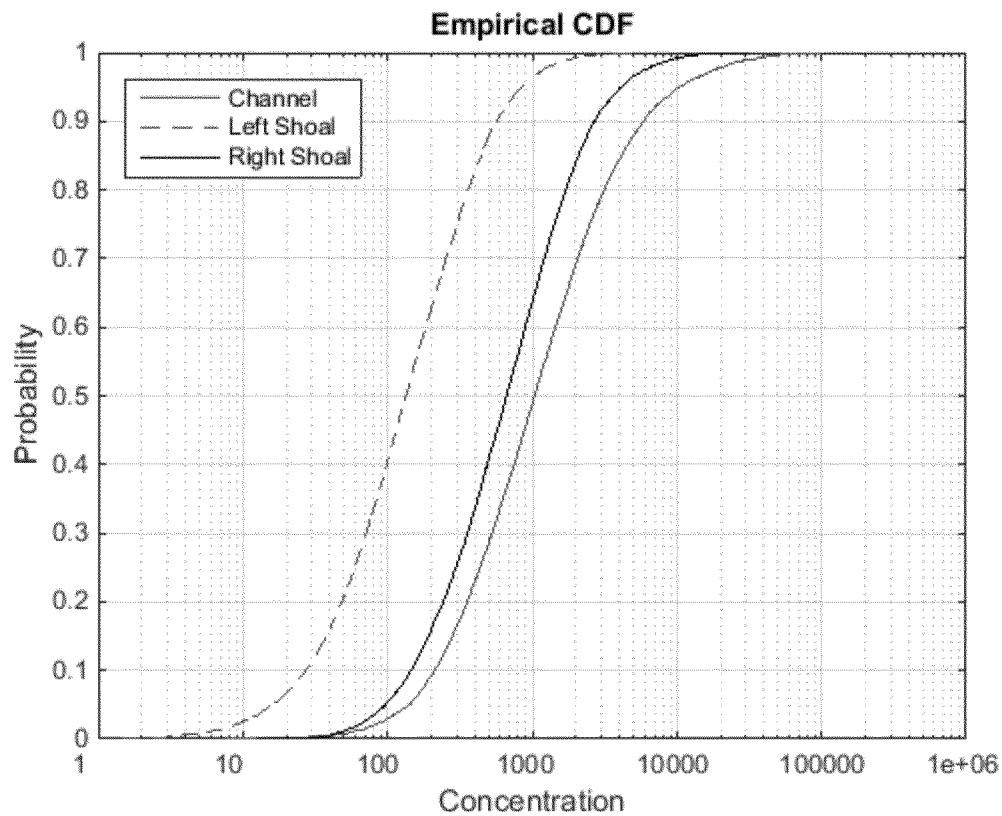
zoneID=[1:8;9:16;17:24];

zone=[kron(zoneID(1,:),ones(20,64));
      kron(zoneID(2,:),ones(24,64))
      kron(zoneID(3,:),ones(20,64))];

% Area of decision units

```

```
duArea=[ones(8,1)*20*64  
        ones(8,1)*24*64  
        ones(8,1)*20*64];  
  
DUX=[1 64 64 1]';  
DUy=[1 21 45  
      1 21 45  
      20 44 64  
      20 44 64];  
  
% Initialize collectors for simulated SWAC values  
SWAC_predicted=[];  
SWAC_actual=[];  
  
Footprint_predicted=[];  
Footprint_actual=[];  
  
FalseNegativeRate=[];  
FalsePositiverate=[];
```



	Left Shoal	Channel	Right Shoal
Geometric Mean	1081	129	649
Mean	2780	249	1221
75th Percentile	2494	301	1427
Standard Deviation	6173	354	1784

Plan View Maps Illustrating Synthetic Surfaces

Synthetic distributions are lognormally distributed with three longitudinal zones representing shoals and channel configuration. The areas are subdivided into 24 decision units which are analogous to the Thiessen polygons that represent "decision areas" in the methods commonly deployed at sediment sites. Each area contains a single sample which is used to decide which areas should be counted toward remediation and which should be left behind under the assumption that they contain contaminant concentrations that are below the selected RAL. Each decision unit contains a number representing the average concentration within that particular DU. In general, when a DU is selected for remediation, it is the average concentration that is removed in the field, as opposed to the individual sample value as assumed in the hilltopping calculations. This assumption is expected to result in overstatement of the benefit of the remedial option under consideration. This simulation checks this expectation and also provides an estimate of the magnitude of the resulting bias.

```
figure;
set(gcf,'position',[106 171 1220 465]);
pcolor(log(Z)), shading('flat');
axis('equal');
hh=colorbar('horiz');

set(hh,'ticks',log([ 0.1 1 10 100 1000 10000 100000]));
set(hh,'ticklabels',[ 0.1 1 10 100 1000 10000 100000 ]);

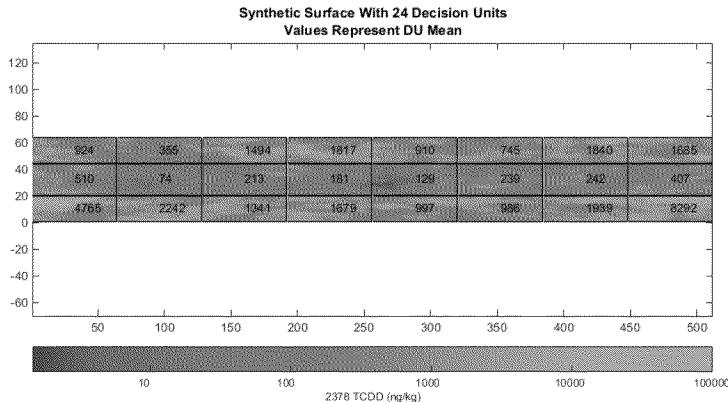
set(get(hh,'label'),'string', '2378 TCDD (ng/kg)');
hold on

CurrentCell=0;
for i=1:3
    for j=1:8
        CurrentCell=CurrentCell+1;
        xtmp=DUX+(j-1)*64;
        ytmp=DUy(:,i);

        Centroid.x=mean(xtmp);
        Centroid.y=mean(ytmp);

        plot(DUX+(j-1)*64,DUy(:,i),'k-');
        hold on;
        h(CurrentCell)=text(Centroid.x,Centroid.y,[num2str(round(mean(Z(zone==Curr
end
```

```
end
title({'Synthetic Surface With 24 Decision Units','Values Represent DU Mean'});
```



Begin Simulation Loop

Multiple synthetic realizations are generated and subjected to the hilltopping algorithm as it is generally applied in RI/FS calculations. The predicted post remedial SWAC is compared with the actual value based on the averages of the synthetic surfaces within each DU.

```
output.XI=XI(:);
output.YI=YI(:);
output. StripID=StripID(:);

for isim=1:n sim
    [X,S]=simfast1(2*n1,n2,dx1,dx2,param,theta,a,b,s2,model);
    X=X(1:n1,1:n2); % Select the top half of the simulated realization to
    % artifacts of periodicity.

    % Add mean normal simulation and exponentiate to produce
    % lognormally distributed surface.
    if UseReportSettings ==1
        Z=exp(X+M);
    else
        Z=exp(X+M+nugget*random('Normal',0,sqrt(s2),size(X)));
    end

    % Store output simulation in MATLAB data structure
    output.simulation(isim).Z=Z(:);

    % Calculate zone summary statistics
    [m,g]=grpstats(Z(:,zone),{'mean','gname'}) ;
    % Zones run from 1 to 24 representing decision units
```

Draw a sample from each cell

This step identifies one location at random within each of the 24 decision units. This single value is assumed to characterize the decision unit in subsequent calculations of the post remedial SWAC.

```
nCell=max(zone(:));  
  
for i=1:nCell  
    k=find(zone==i);  
    j=(k(randperm(length(k))));  
    zsamp(i,1)=Z(j(1));  
end  
  
if isim <5 % ==1 ||isim==2
```

Plan View of a few Synthetic Maps

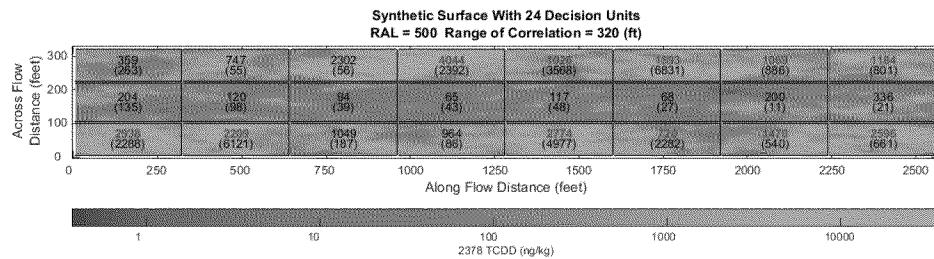
```
figure;  
set(gcf,'position',[106 171 1220 465]);  
pcolor(log(Z)),shading('flat');  
axis('equal');  
hh=colorbar('horiz');  
  
set(hh,'ticks',log([ 0.1 1 10 100 1000 10000 100000]));  
set(hh,'ticklabels',[ 0.1 1 10 100 1000 10000 100000 ]);  
  
set(get(hh,'label'),'string', '2378 TCDD (ng/kg)');  
hold on  
  
CurrentCell=0;  
for i=1:3  
    for j=1:8  
        CurrentCell=CurrentCell+1;  
        xtmp=DUX+(j-1)*64;  
        ytmp=DUY(:,i);  
  
        Centroid.x=mean(xtmp);  
        Centroid.y=mean(ytmp);  
  
        plot(DUX+(j-1)*64,DUY(:,i),'k-');  
        hold on;  
        h(CurrentCell)=text(Centroid.x,Centroid.y+3,[num2str(round(mean(Z(  
text(Centroid.x,Centroid.y-3,[ '*' num2str(round(zsamp(CurrentCell)  
if zsamp(CurrentCell) > RAL;  
        set(h(CurrentCell),'color','r','fontweight','bold');  
    end  
end  
end  
set(gca,'xlim',[-1,514]);  
set(gca,'ylim',[-1,66]);  
xlabel('Along Flow Distance (feet)')  
ylabel({'Across Flow', 'Distance (feet)'})  
set(gca,'xticklabel',5*get(gca,'xtick'))  
set(gca,'yticklabel',5*get(gca,'ytick'))  
title({'Synthetic Surface With 24 Decision Units', ['RAL = ', num2str(RAL)  
  
txt={[ 'Notes: '],...  
}})
```

Hilltop Simulation:
Passaic River RI/FS

[*1) Decision unit average represented by top number: ', num2str(round

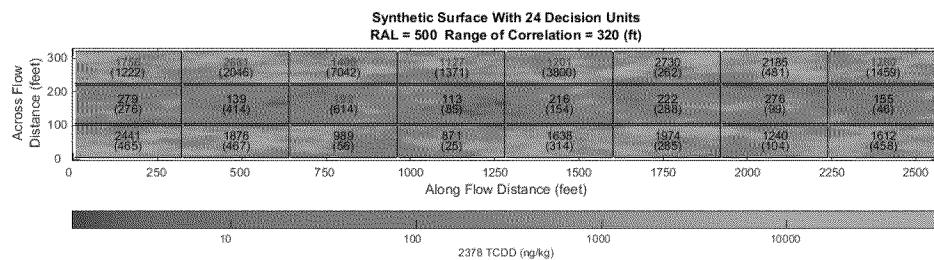
[*2) Single sample value shown in parentheses: (' , num2str(round(zsam

[*3) Red text indicates cells identified for removal because the sampl
ht=annotation('textbox',[0.05,0.05,0.9,0.2],'string',txt,'linestyle','none



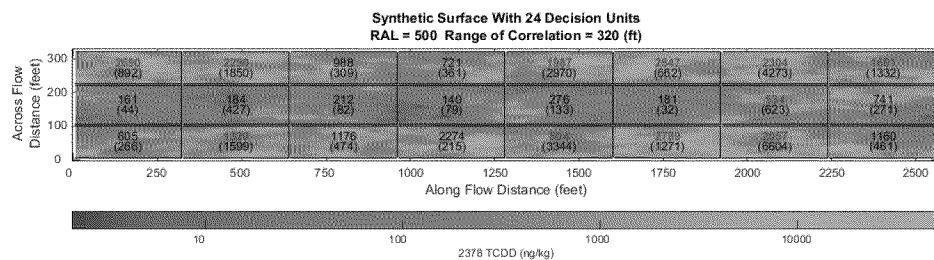
Notes:

- 1) Decision unit average represented by top number: 2938
- 2) Single sample value shown in parentheses: (2288)
- 3) Red text indicates cells identified for removal because the sample value exceeds the RAL = 500



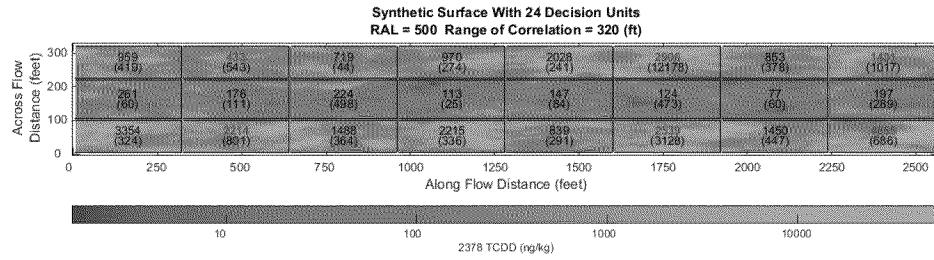
Notes:

- 1) Decision unit average represented by top number: 2441
- 2) Single sample value shown in parentheses: (465)
- 3) Red text indicates cells identified for removal because the sample value exceeds the RAL = 500



Notes:

- 1) Decision unit average represented by top number: 605
- 2) Single sample value shown in parentheses: (268)
- 3) Red text indicates cells identified for removal because the sample value exceeds the RAL = 500



Notes:
 1) Decision unit average represented by top number: 3354
 2) Single sample value shown in parentheses: (324)
 3) Red text indicates cells identified for removal because the sample value exceeds the RAL = 500

Actual vs Predicted Surface

Actual average DU concentrations are plotted against the individual sample values generated from each DU to evaluate the tendency to over or under state the DU concentration when a single sample is drawn from a right skewed distribution. Generally the mean of a log-normal distribution is close to the 60th to 70th percentile of the distribution, so most samples are expected to fall below the true mean. This differs from a symmetric distribution where each sample is equally likely to be above or below the mean.

For DUs with mean concentrations exceeding the RAL, this approximately 65% likelihood that a single sample is less than the global mean implies that at the local DU level there is an increased (but perhaps not 65%) likelihood of misclassifying an offending DU as a non-target DU. These false negative mistakes tend to exacerbate the already understated concentrations in the group not targeted for remediation.

These characteristics are not guaranteed to play out systematically in every effort to quantify the SWAC vs RAL relationship. Generally the effects can be seen in the following figures by noticing that the actual means tend to be higher than predicted values in the non-target group. Conversely, the actual means in the target group tend to be lower (below the 1 to 1 line) than the forecasted values in the target groups. Taken together the effects of the remedy tend to be overstated. This simulation compiles the frequency with which the remedial effectiveness is over vs understated based on many repeated trials applied to synthetic contaminant distributions.

```

figure
set(gcf,'position',[45     141     830     552]);
h1=plot(zsamp,m,'ko');
set(gca,'yscale','log');
set(gca,'xscale','log');
axis('square');
axis('equal');
hold on;
h2=plot([10,10000],[10,10000],'r--','linewidth',2);
grid on;
ylabel('Actual Decision Unit Mean');
xlabel('Sample Value Per Decision Unit');
title({'Actual vs Predicted Decision Unit Mean',...
    'One Particular Synthetic Map'});
j=find(zsamp>RAL);
h3= plot(mean(zsamp(j)),mean(m(j)),'ks','markerfacecolor','r','markersize'

```

Hilltop Simulation:
Passaic River RI/FS

```
j=find(zsamp<=RAL);
h4=plot(mean(zsamp(j)),mean(m(j)), 'ks', 'markerfacecolor', 'b', 'markersize',
set(gca, 'ylim',[10,10000]);
set(gca, 'xlim',[10,10000]);;

yl=get(gca, 'ylim');
xl=get(gca, 'xlim');

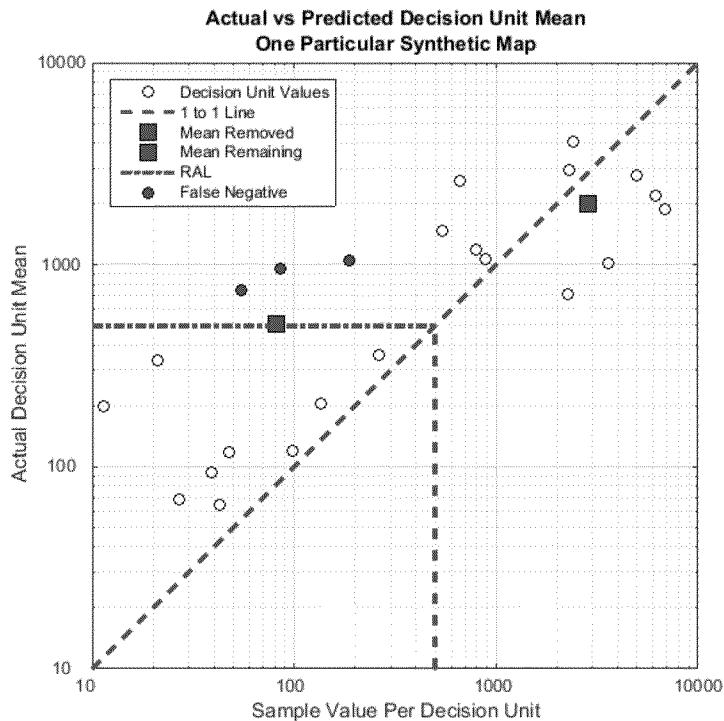
h5=plot([xl(1) RAL], [RAL,RAL], 'g-.','linewidth',2);
plot([RAL,RAL], [yl(1) RAL], 'g--.','linewidth',2);

% False negative
j=find(zsamp<RAL & m > RAL);
h6=plot(zsamp(j),m(j), 'ko', 'markerfacecolor', 'r');

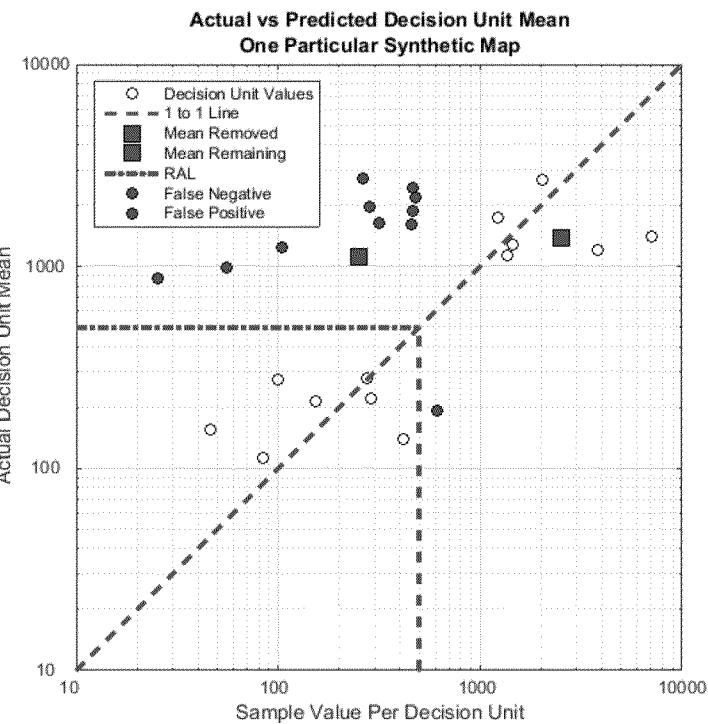
% False Positive
j=find(zsamp>RAL & m < RAL);
h7=plot(zsamp(j),m(j), 'ko', 'markerfacecolor', 'b');

h=legend([h1,h2,h3,h4,h5,h6,h7],{'Decision Unit Values','1 to 1 Line',...
'Mean Removed',...
'Mean Remaining',...
'RAL','False Negative','False Positive'},'location','NorthWest');
set(gca, 'xticklabel',(get(gca,'xtick'))');
set(gca, 'yticklabel',(get(gca,'ytick'))');
```

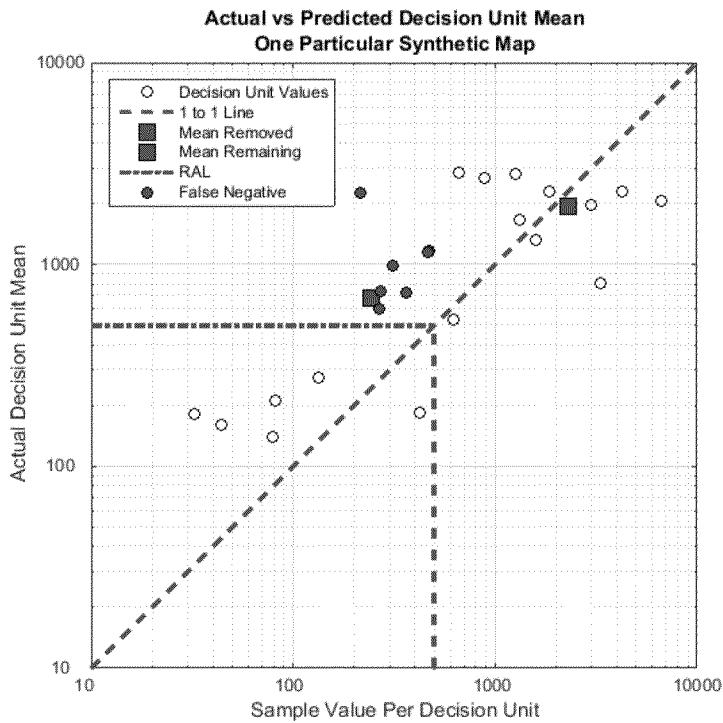
Warning: Ignoring extra legend entries.

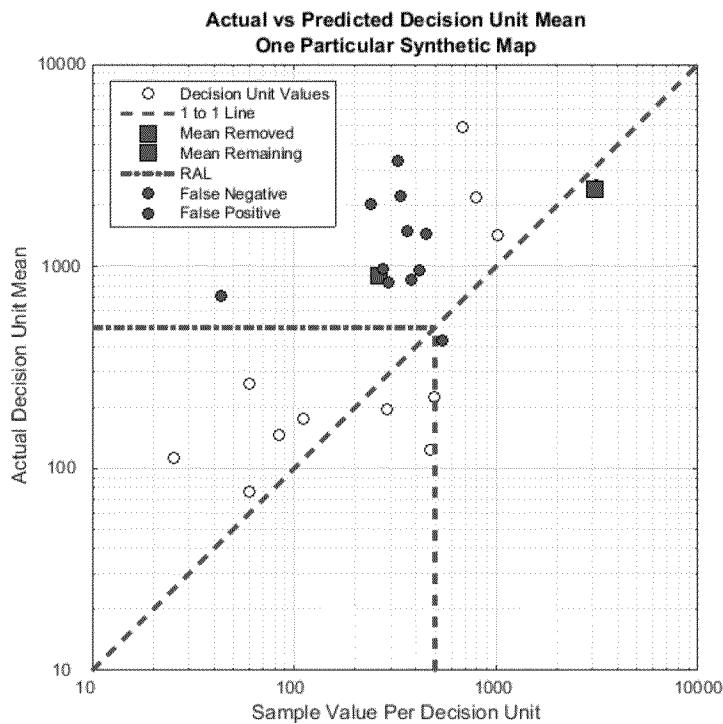


Hilltop Simulation:
Passaic River RI/FS



Warning: Ignoring extra legend entries.





end

Calculate Actual and Predicted Post Removal SWAC

The actual post remedial SWAC is given by averaging the synthetic surface concentrations with cells identified for remediation replaced with zeros. In practice we might replace with a non-zero value to represent post remedial residuals or recontamination. Setting to zero as opposed to an alternative non-zero value has little or no influence on the comparison of predicted and actual SWAC.

```

jin=find(zsamp>RAL); % jin sample nodes are in the remedial footprint
jout=find(zsamp <=RAL); % jout grid nodes are outside the

if UseReportSettings ==1
    SWAC_predicted=[SWAC_predicted;mean([zeros(size(jin));zsamp(jout)])];
    SWAC_actual=[SWAC_actual;mean([zeros(size(jin));m(jout)])];
else

```

Area weighted average -- Revised to accomodate unequal decision unit sizes

```

SampTmp=zsamp;
SampTmp(jin)=0;
SWAC_predicted=[SWAC_predicted; sum(SampTmp.*duArea)/sum(duArea)];

```

```
TrueTmp=m;
TrueTmp(jin)=0;
SWAC_actual=[SWAC_actual; sum(TrueTmp.*duArea)/sum(duArea) ];

end

Footprint_predicted=[Footprint_predicted; sum(duArea(jin))];

jTrueFP=find(m>RAL);
Footprint_actual=[Footprint_actual; sum(duArea(jTrueFP))];

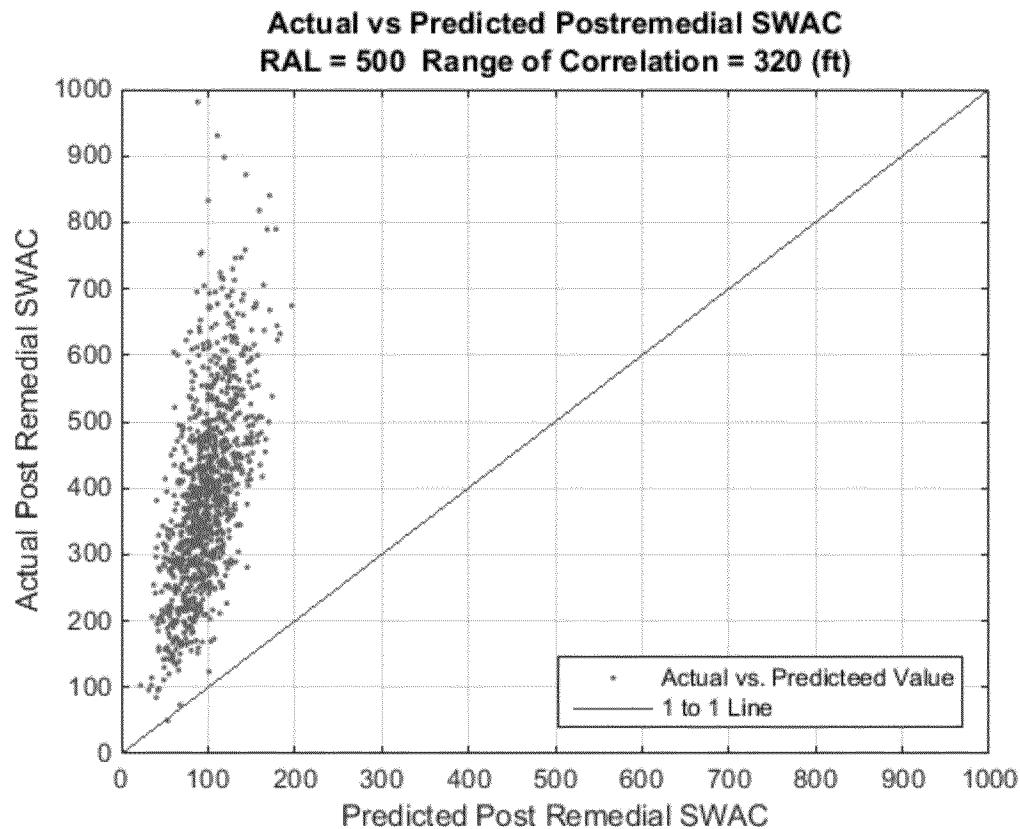
end
```

Actual vs Predicted SWAC

If the hilltopping algorithm is unbiased the predicted and actual SWACs should plot on the 1 to 1 line

```
figure
plot(SWAC_predicted,SWAC_actual,'.');
hold on;

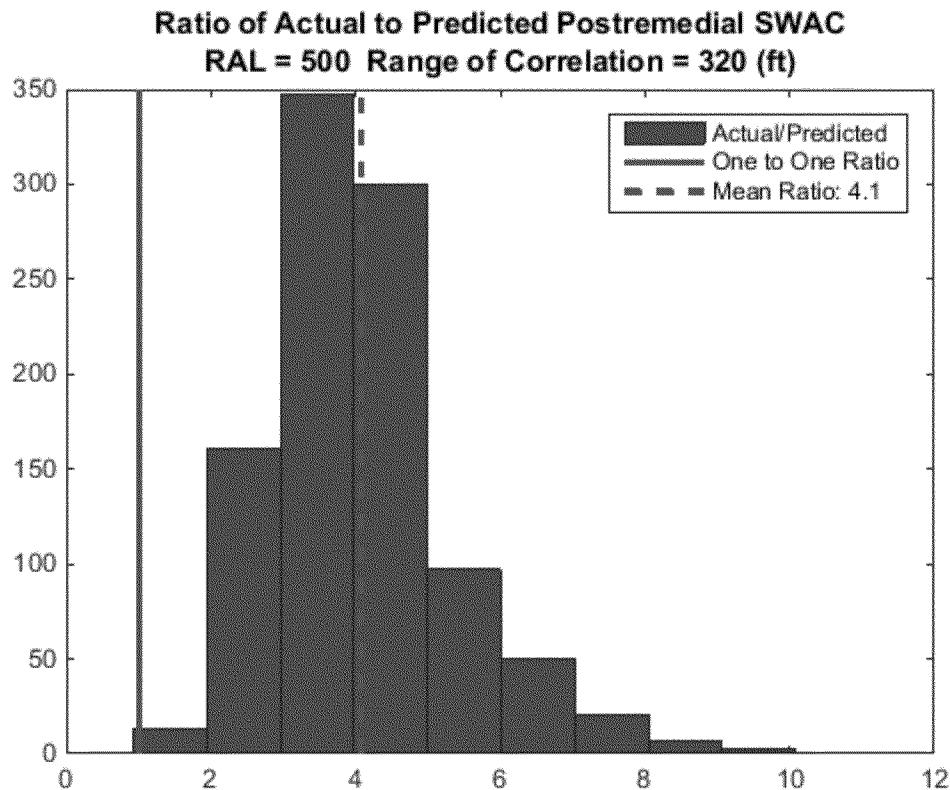
yl=get(gca,'ylim');
yl(1)=0;
set(gca,'ylim',yl);
set(gca,'xlim',yl);
plot(yl,yl,'r-');
grid on;
ylabel('Actual Post Remedial SWAC');
xlabel('Predicted Post Remedial SWAC');
title({'Actual vs Predicted Postremedial SWAC', ['RAL = ', num2str(RAL), ' Range'],
legend({'Actual vs. Predicted Value', '1 to 1 Line'}, 'Location', 'SouthEast')}
```



Distribution of Ratio of Actual to Predicted SWAC

If the hilltopping algorithm is unbiased the ratio of actual to predicted should be centered at 1.0. Notice in the following graphic, the ratios are nearly always greater than 1.0 indicating that the predicted SWAC based on Hilltopping, understates the post remedial SWAC, or equivalently overstates the benefit of the remedy.

```
figure;
R=SWAC_actual./SWAC_predicted;
hist(R);
hold on;
yl=get(gca,'ylim');
plot([1 1],yl,'r-','linewidth',2);
plot([nanmean(R),nanmean(R)],yl,'g--','linewidth',2);
legend('Actual/Predicted','One to One Ratio',['Mean Ratio: ' num2str(round(nanmean(R)))]);
title({'Ratio of Actual to Predicted Postremedial SWAC',...
['RAL = ', num2str(RAL), ' Range of Correlation = ', num2str(round(1/a*pi))]});
```



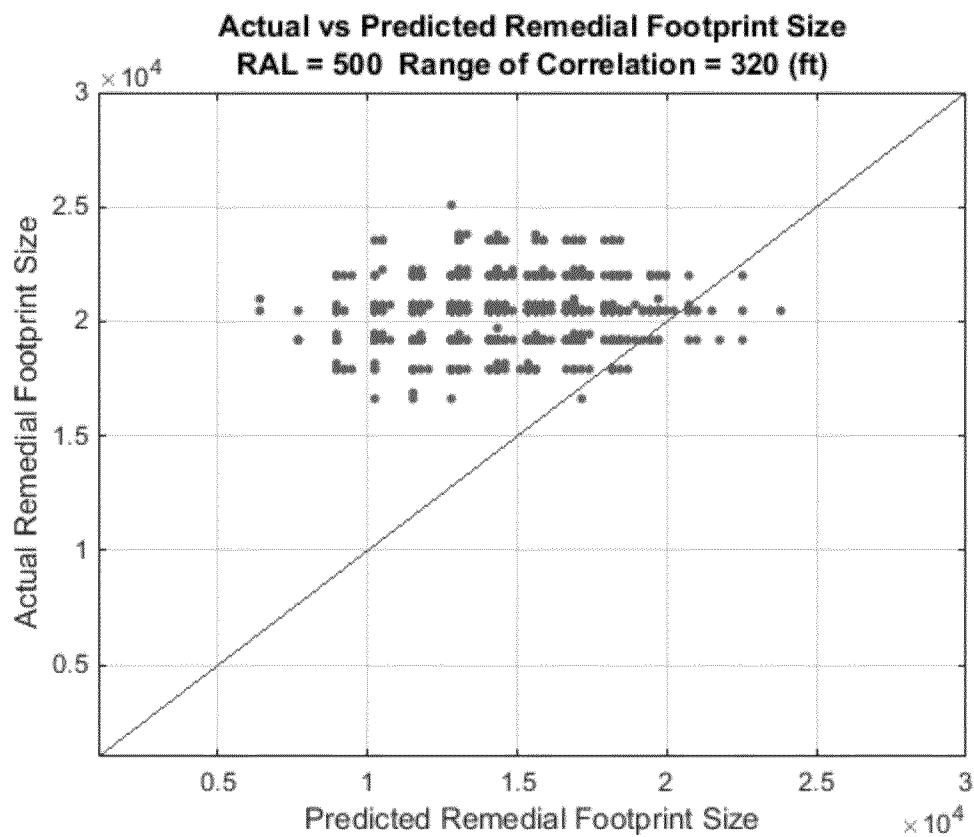
Actual vs Predicted Remedial Footprint Area

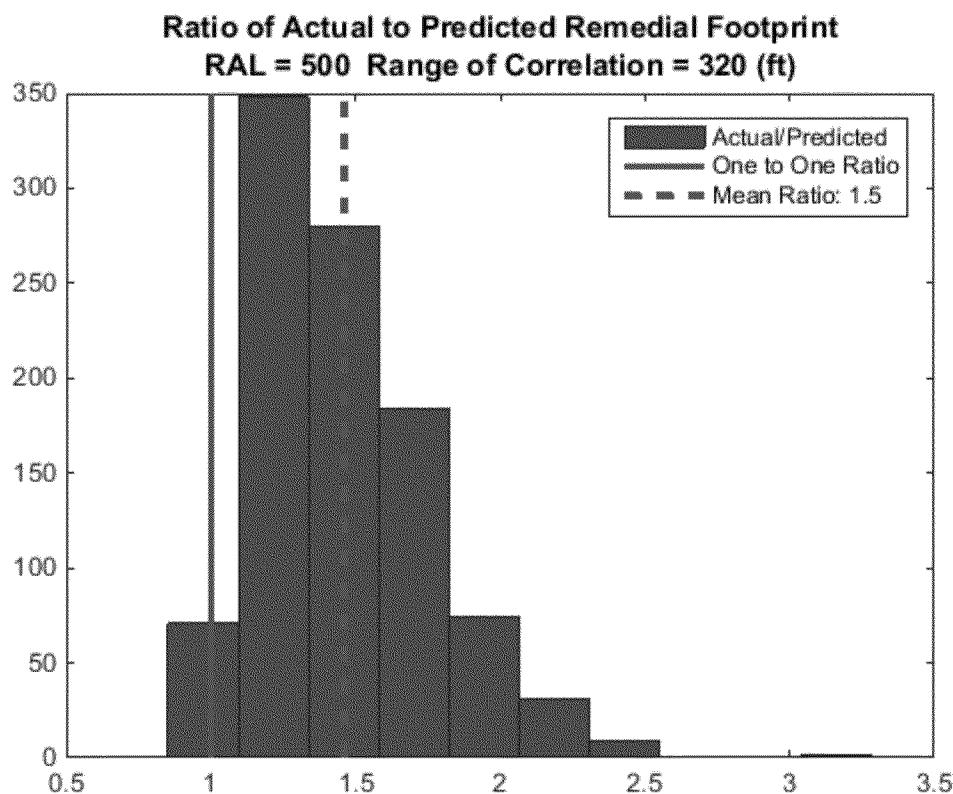
```
figure
plot(Footprint_predicted,Footprint_actual,'.', 'markersize',10)
hold on
xlabel('Predicted Remedial Footprint Size')
ylabel('Actual Remedial Footprint Size')
% yl=get(gca,'ylim');
% yl(1)=0;
set(gca,'ylim',[1000,30000]);
set(gca,'xlim',[1000,30000]);
plot([1000,30000],[1000,30000],'r-');
grid on;
title({'Actual vs Predicted Remedial Footprint Size',...
    ['RAL = ', num2str(RAL), ' Range of Correlation = ', num2str(round(1/a*pi)]});
% set(gca,'yscale','log')
% set(gca,'xscale','log')

figure;
Rfp=Footprint_actual./Footprint_predicted;
hist(Rfp);
hold on;
yl=get(gca,'ylim');
plot([1 1],yl,'r-','linewidth',2);
```

Hilltop Simulation:
Passaic River RI/FS

```
plot([nanmean(Rfp),nanmean(Rfp)],yl,'g--','linewidth',2);
legend('Actual/Predicted','One to One Ratio',[ 'Mean Ratio: ' num2str(round(nanmean(Rfp)))...
title({'Ratio of Actual to Predicted Remedial Footprint',...
['RAL = ', num2str(RAL), ' Range of Correlation = ' num2str(round(1/a*pi))]);
```





Published with MATLAB® R2014b